Metafounder approach for single-step genomic evaluations of Red Dairy cattle

Kudinov, A. A.

2020-07

Kudinov , A A , Mantysaari , E A , Aamand , G P , Uimari , P & Stranden , I 2020 , ' Metafounder approach for single-step genomic evaluations of Red Dairy cattle ' , Journal of Dairy Science , vol. 103 , no. 7 , pp. 6299-6310 . https://doi.org/10.3168/jds.2019-17483

http://hdl.handle.net/10138/329937
https://doi.org/10.3168/jds.2019-17483

Interpretive Summary

Metafounder approach for single-step genomic evaluations of Red Dairy cattle. By Kudinov et al. Change from the multi-step to the single-step genomic prediction approach in routine evaluations is complicated. In this study, we show the advantage of the metafounders approach in the single-step prediction of milk performance in dairy cattle. In addition, we also test the effect of markers selection on creating a metafounders relationship matrix.

Metafounder approach for single-step genomic evaluations of Red Dairy cattle

A.A. Kudinov[*, †], E.A. Mäntysaari[*], G.P. Aamand[‡], P. Uimari[†], and I. Strandén[*]

[*] Natural Resources Institute Finland (Luke), Jokioinen, Finland, FI-31600

[†] Department of Agricultural Science, University of Helsinki, Helsinki, Finland, FI-00014

[‡] Nordic Cattle Genetic Evaluation, Aarhus, Denmark, DK-8200

Corresponding author: Andrei Kudinov

e-mail: andrei.kudinov@luke.fi

ABSTRACT

Single-step genomic BLUP (**ssGBLUP**) is a powerful approach for breeding value prediction in populations with a limited number of genotyped animals. However, conflicting genomic (**G**) and pedigree ($_{22}$) relationship matrices complicate the implementation of ssGBLUP into practice. The metafounder (**MF**) approach is a recently proposed solution for this problem and has been successfully used on simulated and real multi-breed pig data. Advantages of the method are easily seen across breed evaluations, where pedigrees are traced to several pure breeds, which are thereafter used as MF. Application of the MF method to ruminants is complicated due to multi-breed pedigree structures and the inability to transmit existing unknown parent groups (**UPG**) to MF. In this study, we apply the MF approach for ssGBLUP evaluation of Finnish Red Dairy cattle treated as a single breed. Relationships among MF were accounted for by a (co)variance matrix (**Γ**) computed using estimated base population allele frequencies. The attained **Γ** was used to calculate a relationship matrix $_{22}$ for the genotyped animals. We tested the influence of SNP selection on the **Γ** matrix by applying a minor allele frequency (**MAF**) threshold ($_{MAF}$) where accepted markers had an MAF ≥ 0.05. Elements in the $_{MAF}$ matrix were slightly lower than in the **Γ** matrix. Correlation between diagonal elements of the genomic and pedigree relationship matrices increased from 0.53 ($_{22}$) to 0.76 ($_{22}$ and $_{22}^{MAF}$). Average diagonal elements of $_{22}$ and $_{22}^{MAF}$ matrices increased to the same level as in the **G** matrix. ssGBLUP breeding values (**GEBV**) were solved using either the original 236 or redefined 8 UPG, or 8 MF computed with or without the MAF threshold. For bulls, the GEBV validation test results for the 8 UPG and 8 MF gave the same adjusted R^2 (0.31) and over-dispersion (0.73, measured by regression coefficient $_1$). No significant R^2 increase was observed in cows. Thus, the MF greatly influenced the pedigree relationship matrices but not the GEBV.

Selection of SNPs according to MAF had a notable effect on the **Γ** matrix and made the $\mathbf{A}_{22}$ and **G** matrices more similar.

Key Words

Genetic groups, single-step genomic BLUP, metafounders, base population.

INTRODUCTION

Single-step genomic BLUP (**ssGBLUP**) is an elegant approach for estimating genomic breeding values (**GEBV**) that uses pedigree ( ) and genomic ( ) relationship matrices (Aguilar et al., 2010; Christensen and Lund, 2010). The approach has two important theoretical assumptions concerning the and matrices: the same scale and equal base population (Christensen, 2012). These assumptions complicate the application of ssGBLUP in dairy cattle breeding. In order to meet the assumptions, several methods have been proposed that make to be like . For example, base population allele frequencies (**AF**) are used (VanRaden, 2008), and elements of are scaled and centered to have on average the same diagonal and off-diagonal elements as in (Vitezica et al., 2011; Christensen et al., 2012). In practice, base population AF are unknown and the matrix is often constructed using AF observed in the genotyped population.

Commercial dairy cattle pedigree can seldom be traced to a genetically homogeneous base population because the pedigree often has a complicated breed structure with unknown parent information (VanRaden, 1992; Sponenberg and Bixby, 2007). To solve the problem of incomplete pedigree, Thompson (1979) and Quaas (1988) developed the concept of phantom parents or unknown parent groups (**UPG**), for animals with unknown parent(s). UPG are typically assigned according to selection pathways and share the same genetics allowing

more accurate estimation of genetic trend in traditional genetic evaluation (Theron et al., 2002). In ssGBLUP, Misztal et al. (2013) observed bias in UPG solutions. The bias increased with an increase in the number of genotyped animals.

The metafounder (**MF**) approach was proposed by Legarra et al. (2015) to achieve compatibility in the pedigree and genomic relationship matrices. The MF approach combines the idea of using AF equal to 0.5 for all markers when calculating the matrix (Christensen, 2012) and assigning unknown parents to MF or pseudo-individuals with self-relationships in the matrix. MF are similar to UPG, but allow a related base population with non-zero inbreeding coefficients. The relationships within and between the MF are modeled by a gamma matrix ( ), which is used in forming the relationship matrix ( ). The matrix may be constructed using an estimated base or observed genotyped population AF (e.g. Legarra et al., 2015; Garcia-Bacciano et al., 2017). However, the matrix may be poorly estimated when certain AF are estimated inaccurately due to the low number of rare alleles. The large number of UPG increases chances that an UPG is associated with a low number of rare allele genotypes.

Legarra et al. (2015) and Garcia-Bacciano et al. (2017) showed the advantage of the MF approach in GEBV estimation using simulated data. Xiang et al. (2017) used the MF method for ssGBLUP evaluation in the crossbreed performance in pigs. According to their results, the MF approach successfully combined two breeds in a GEBV evaluation. Pig evaluations clearly focus on the youngest generation and, thus, fewer UPG are needed than in dairy cattle (Arnold et al., 1992). MF approach studies have mostly focused on crossbred and admixture populations (Bradford et al., 2019; van Grevenhof et al., 2019) because the approach may help with implementing ssGBLUP for complicated pedigree populations such as in pigs and poultry. However, implementing the MF approach for dairy cattle may be challenging

because of the frequently large number of UPG. The few published studies have used simulated dairy cattle data to estimate the matrix and its influence on ssGBLUP (Garcia-Bacciano et al., 2017; Bradford et al., 2019), but had only a few MF.

We used the MF approach in the ssGBLUP evaluation of 305-d milk production in Finnish Red dairy cattle. We present two approaches to estimate the matrix, using different numbers of markers. We compared values in the two matrices. The effect of various matrices is shown using model validation statistics from ssGBLUP evaluations having either UPG or MF.

MATERIALS AND METHODS

ssGBLUP models

The joint relationship matrix of genotyped and non-genotyped animals in ssGBLUP is commonly denoted as **H** (Aguilar et al., 2010; Christensen and Lund, 2010). The $^{-1}$ matrix needed in the mixed model equations of ssGBLUP is

$$^{-1} = \quad ^{-1} + (\quad ^{-1} - \quad ^{-1}_{22}),$$

where is the full pedigree relationship matrix, is the genomic relationship matrix, and $_{22}$ is a pedigree relationship matrix of the genotyped animals.

Single step with UPG in A. Mean genetic levels of animals with missing parental information were modeled using pedigree-based UPG proposed by Quaas and Pollack (1981). In the UPG model, unknown parents are assumed to be unrelated and completely outbred. UPG effects in the model only account for possible non-zero expectations in the breeding values of parent groups. There are alternative ways to account for UPG in forming $^{-1}$. The standard way is to replace the original $^{-1}$ matrix with an augmented one, where the UPG are included as

112 "phantom parents" (Westell et al., 1988). Matilainen et al. (2018), following Misztal et al.
113 (2013), formed the $^{-1}$ matrix without groups, and, thereafter, included the UPG via so-
114 called QP transformation (Quaas and Pollack, 1981) into the final augmented $^{-1}$. However,
115 Masuda et al. (2019) recommended omitting the terms involving $^{-1}$ in the UPG coefficient
116 part of the augmented $^{-1}$ matrix. In our UPG models, the genomic relationship matrix was
117 constructed using VanRaden (2008) method 1 ( ) , where base population AF were used
118 to center and scale the marker data. Base population AF were estimated with the GLS model
119 (McPeek et al., 2004) using the Bpop v. 0.30 program (Strandén and Mäntysaari, 2019),
120 which is based on the computational approach described in Strandén et al. (2017). The
121 genomic information was assumed to account for 90% of the variation in breeding values, i.e.
122 the polygenic proportion was 10%. This was attained using a modified matrix obtained by
123 averaging original and $_{22}$ matrices with weights of 0.9 and 0.1, respectively.

124 ***Single step with metafounders***. In the MF approach, the $^{-1}$ matrix is replaced by a
125 modified $(\ )^{-1}$ matrix described by Legarra et al. (2015) and Christensen et al. (2015) as

126 $$(\ )^{-1} = (\ )^{-1} + (\ \ \ ^{-1} - (\ _{22})^{-1}),$$

127 where $= (1 -)\ + \ _{22}$, w is the proportion of genetic variance not explained by
128 the markers, $= (\ _{101}\ \ ^{\prime}_{101})\frac{2}{}$, $_{101}$ is an n by m marker matrix with genotypes coded by
129 {-1,0,1}, is the number of SNP markers, n is the number of genotyped animals, is
130 pedigree relationship matrix formed with a matrix, and $_{22}$ is a submatrix of for the
131 genotyped animals. We used a 10% polygenic proportion, i.e. w = 0.1, as in Garcia-Baccino
132 et al. (2017). The variance covariance structure of the MF can be estimated by $= 8\ \ (\ )$,
133 as presented in the Appendix of Christensen et al. (2015), where is an m by r matrix of AF
134 and r is the number of MF.

135 ***Test data and model validation***

136 We used Red Dairy Cattle (**RDC**) milk production data provided by Nordic Cattle Genetic
137 Evaluations (**NAV**). The data sample was extracted from the NAV production evaluation
138 database by including all cows from 426 Finnish herds with at least 10 genotyped cows. This
139 gave 112,479 cows with first-lactation 305-d milk production records produced during 1988–
140 2018. The pedigree included 226,012 animals born in 1960–2016 consisting of 86% RDC,
141 12% Holstein (**HOL**), 2% Finn cattle (**FIN**, an indigenous Finnish cattle population), and a
142 total of 1% of other breeds (Red Holstein, Jersey, Brown Swiss etc.). There were 236 UPG
143 which were based on selection path, birth year, and population of origin. These UPG
144 definitions were the same as those used in the Nordic TD evaluations in November 2018
145 (Lidauer et al., 2015) and were provided by NAV.

146 Genotypes were available for 19,757 animals (3,571 bulls and 16,186 cows), which either
147 had observations or were in the pedigree of the animals with observations. Bulls were
148 genotyped using Illumina Bovine SNP50 Bead Chip (Illumina, San Diego, USA) and the
149 cows using a lower-density EuroG 10k chip (http://www.eurogenomics.com/) that had been
150 imputed to the 50K density by NAV. There were 46,914 markers from the 29 bovine
151 autosomes available for the analysis.

152 Cow and bull validation data sets were created by removing milk production records for
153 either the last year or for four of the previous production years, respectively, as in Gao et al.
154 (2018) and Mäntysaari et al. (2010). We included 101 and 3,551 genotyped test bulls and
155 cows, respectively. Daughter yield deviations (**DYD**) and yield deviations (**YD**) were attained
156 using the full data and an animal model by the MiX99 software (Strandén and Lidauer,
157 1999), as in Gao et al. (2018). The calculated DYD and YD were used for bulls and cows,
158 respectively, in validation regression models $( ) = {}_0 + {}_1$, with weights for

159 the DYD observations. The weight for DYD was /(+), where is $(4 - h^2) / h^2$,
160 h^2 is heritability, and EDC is the bull's effective daughter contributions
161 (https://interbull.org/ib/cop_appendix4) in evaluation with the full data set. To attain adjusted
162 validation reliability, we divided the model coefficient of determination ( ) by the average
163 weight. The regression coefficient b_1 for the bulls was multiplied by two because DYD only
164 represents the sire effect. All the analyses used h^2 of 0.44, which is a parameter derived from
165 the NAV milk production test day model for 305-d milk yield.

166 ***Unknown parents and metafounders***

167 Eight groups were defined according to the full pedigree structure and replaced the original
168 236 UPG. We included six groups for RDC (birth years <1971, 1971–1980, 1981–1990,
169 1991–2000, 2001–2010, 2011–2016), a HOL group, and a group for the other breeds. These
170 eight groups were treated as UPG or MF. In the MF approach, the base population AF, used
171 to calculate the matrix, were estimated using a GLS approach. The GLS model was =
172 + , where is an *n* by 1 vector of marker *i* genotypes, is an *n* by 8 matrix, the rows
173 of which sum up to 1, and that assigns individuals to fractions of MF, is an 8 by 1 vector
174 of group means, and ~ (, $_{22}\sigma^2$) where $_{22}$ was the pedigree relationship matrix for the
175 genotyped animals and σ^2 is the common variance. In allele frequency estimation, the
176 common variance need not be known (e.g. Garcia-Baccino et al., 2017). Estimated base
177 population AF for the MF are $\hat{} = \frac{1}{2}\hat{}$ for each marker $= 1, \ldots,$.

178 To estimate AF for the MF in the GLS model, the $_{22}$ matrix was based on a truncated
179 pedigree, where one parent generation at most was accepted to the genotyped animals. The
180 pedigree truncation guaranteed that the young genotyped animals would contribute to the
181 recent birth year MF and not to the old birth year MF. In addition, the truncation used more
182 genomic information than the full pedigree because genotyped animals had less genotyped

ancestors but instead a young birth year MF. It can be proven that the GLS method will ignore genotype of an animal whose both parents are genotyped and the animal is not an ancestor to a genotyped animal.

The eight columns of base population AF in the matrix were used to estimate the variance covariance structure of the eight MF or the matrix, = 8 ( ). The effect of minor allele frequencies (**MAF**) on the MF covariances were tested by creating two alternative matrices. In the first scenario, the full matrix was used to calculate the matrix, denoted $_8$. In the other scenario, denoted $_{8MAF}$, only those markers with MAF greater or equal to 0.05 in all RDC cattle MF were included in the **P** matrix. The MAF requirement eliminated 3,783 markers and left 43,131 markers that were used to calculate the $_{8MAF}$ matrix.

ssGBLUP computation

All ssGBLUP calculations used the full pedigree with 226,012 animals and genomic relationship matrices (or) for the 19,757 animals. For the ssGBLUP with MF, the augmented additive relationship matrix of genotyped animals ($_{22}$) was calculated using the modified RelaX2 v. 1.83 program (Strandén and Vuori, 2006). The ($^{-1}$ – $^{-1}_{22}$) and ($^{-1}$ – ($_{22}$)$^{-1}$) matrices were calculated using the HGinv v. 0.87 program (Strandén and Mäntysaari, 2018). The latest MiX99 v. 17.1107 (Strandén and Lidauer, 1999) was used to solve the GEBV using the four ssGBLUP models. Two of the evaluations were UPG models with either 236 UPG ($ssGBLUP_{236UPG}$) or 8 UPG ($ssGBLUP_{8UPG}$) in **A**. UPG were treated as random by adding the inverse of genetic variance to the diagonal of group equations in the mixed model equations. The other two ssGBLUP evaluations were MF models that had eight MF, and the pedigree relationship matrices were based on $_8$ ($ssGBLUP_{8}$) or $_{8MAF}$ ($ssGBLUP_{8MAF}$). Genetic variance parameters from the model with unrelated founders were used to estimate corresponding parameters for the model

with MF. The variance of breeding values in base population descending from MF ($^{2}_{,}$) in $ssGBLUP_{8}$ and $ssGBLUP_{8MAF}$ models were calculated using the scaling parameter , i.e., $^{2}_{,} = {}^{2}/$, where $= (1 + tr( )/(2) - ' / {}^{2})$ and tr( ) is the sum of diagonal elements of the matrix (Legarra et al. 2015).

Comparisons

Two traditional ssGBLUP evaluations were computed using different numbers of UPG, and two MF-based ssGBLUP evaluations were computed using different matrices and inbreeding coefficients. We present the two matrices such that the direct effect of the MAF threshold marker selection is seen in elements of the matrices. The MF approach is expected to give more similar pedigree and genomic relationship matrices than the traditional pedigree and genomic relationship matrices. In addition, the off-diagonal elements in the pedigree relationship matrix by the MF approach are expected to be higher than in the traditional pedigree relationship matrix. We assessed differences in the diagonal elements (related to the definition of inbreeding) and off-diagonals (related to relatedness) of $_{22}$, $_{22}$, , and by correlations and mean differences between these matrices. To identify differences in trends of diagonals to the pedigree and genomic matrices (that are related to breeding selection and changes in inbreeding), average diagonal elements of $_{22}$, , and $_{22}$ were plotted by birth year.

The two UPG definitions and two MF matrices gave four sets of ssGBLUP predictions. Validation tests used GEBV from the ssGBLUP evaluations separately from the groups of genotyped bulls and cows. Approximately 80% of bulls born in 1990 to 2014 were genotyped. Thus, differences between the ssGBLUP models may be largest in the genetic trends of the bulls. Averages and standard deviation of selected bull GEBVs by birth year were plotted for comparison purposes. The bulls selected for plotting had at least 10

231 daughters each. Average cow GEBVs by birth year were plotted using GEBVs from all cows
232 to illustrate the genetic trend in the general population.

233 **RESULTS AND DISCUSSION**

234 ***Elements of , , , , and***

235 Table 1 has elements of the $_{8}$ and $_{8MAF}$ matrices. Elements of the $_{8MAF}$ matrix were
236 slightly lower than corresponding elements in the $_{8}$ matrix. All diagonal elements in the
237 matrices were less than one, which corresponds to negative inbreeding of MF (Table 2)
238 calculated as F = ɣ - 1, where ɣ is the relationship across gametes (diagonal element of).
239 All elements in the calculated $_{8}$ and $_{8MAF}$ matrices were from 0.452 to 0.797.

240 Because the MF were partially formed by breed, the greater than zero off-diagonal elements
241 suggest shared genetics between breeds. Average mean relationship between the RDC and
242 HOL metafounders was 0.564 and 0.473 in $_{8}$ and $_{8MAF}$, respectively. Off-diagonal
243 elements of the matrix between Holstein and Jersey cattle in Legarra et al. (2015) was 0.48,
244 which is close to the value we obtained in $_{8MAF}$. They calculated the matrix using
245 published statistics in VanRaden et al. (2011), which included only SNP markers with MAF
246 ≥ 0.05 (Wiggans et al., 2009). The self-relationships in the HOL and RDC metafounders in
247 our study were also comparable to 0.55 presented for the HOL and Jersey breeds in Legarra
248 et al. (2015). In our study, an exception to this was the RDC < 1970 group, which had a
249 diagonal value of 0.618 and 0.719 in $_{8MAF}$ and $_{8}$, respectively. The larger diagonal value in
250 the oldest RDC group may be due to changes in the Finnish RDC breeding program. Before
251 1970, breeding in the RDC group was mostly limited to Ayrshire cattle with only a low
252 number of imported animals. After 1970, importation began changing the population to more
253 resemble a mixed Nordic RDC breed. Diagonal elements in the group of other breeds were

254 high in both of the matrices (0.740 and 0.797). This may be due to the influence of Finn
255 Cattle having only a small number of animals, which may produce unreliable AF estimates.

256 Table 3 shows correlations between (off-)diagonal elements of $_{22}$, , , $^{8}_{22}$, and
257 $^{8MAF}_{22}$ matrices. Constructing $_{22}$ using $_{8}$ and $_{8MAF}$ increased the correlation between the
258 diagonal elements of and $_{22}$ from 0.66 to 0.76. The diagonal element correlation
259 between elements of $^{8MAF}_{22}$ and $_{22}$ was higher (0.84) than between $^{8}_{22}$ and $_{22}$ (0.81).
260 The correlation between diagonal elements of and $_{22}$ decreased from 0.53 to 0.33 and
261 0.37 for $^{8MAF}_{22}$ and $_{22}$, respectively. Despite the high correlation of 0.99 between the
262 diagonal elements of $^{8}_{22}$ and $^{8MAF}_{22}$, average diagonal elements by the birth year of an
263 animal (Figure 1) were at a higher level for $^{8}_{22}$ than for $^{8MAF}_{22}$ or . Average diagonal
264 elements for both augmented matrices ($^{8}_{22}$ and $^{8MAF}_{22}$) were at the same level as , i.e.,
265 from 1.30 to 1.38, while the average diagonals of $_{22}$ and were in range from 0.98 to
266 1.08. According to the summary statistics in Table 4, values for the off-diagonal elements of
267 the pedigree relationship matrix $_{22}$ increased when using to make $_{22}$. Hence, all
268 elements in the , $^{8}_{22}$, and $^{8MAF}_{22}$ matrices were higher on average than those in the $_{22}$
269 and matrices. Interestingly, both the diagonal and off-diagonal element mean,
270 minimum, and maximum values of and $^{8MAF}_{22}$ agreed very well.

271 Average inbreeding coefficients in the $_{22}$ and matrices were 0.02 and 0.31,
272 respectively. This difference of 0.29 was close to the 0.272 reported in VanRaden et al.
273 (2011) for HOL cattle (0.056 for $_{22}$ and 0.328 for). The average inbreeding coefficient
274 increased from 0.02 in $_{22}$ to 0.34 and 0.29 in $^{8}_{22}$ and $^{8MAF}_{22}$, respectively. Following
275 Legarra et al. (2015), a diagonal element value less than one in the matrix means a negative
276 individual inbreeding coefficient for MF. In all RDC MF, all elements of diag( )-1 ranged

277 from -0.38 to -0.43. We observed the highest self-relationships and corresponding MF
278 inbreeding coefficients in the other breed group, which could be explained by the relatively
279 closed small-scale selection program for FinnCattle.

280 Use of the matrix to make the pedigree-based relationship matrix $_{22}^{8}$ or $_{22}^{8MAF}$ increased
281 the correlation between elements of the pedigree and genomic relationship matrices when
282 compared to the correlation between traditionally formed matrices (and $_{22}$).
283 Correlation between diagonal elements of $_{22}^{8}$ and , as well as between $_{22}^{8MAF}$ and ,
284 was 0.76, which is higher than the correlation of 0.53 between the diagonal elements of
285 and $_{22}$. Correlation between the off-diagonal elements of $_{22}^{8}$ ($_{22}^{8MAF}$) and was
286 0.91, which is a bit higher than the same correlation (0.89) between and $_{22}$. Thus,
287 using the matrix to form the relationship matrix lifted the diagonal elements of $_{22}$ matrix
288 to the same level as in the matrix (Figure 1).

289 The average diagonal of the $_{22}^{8}$ matrix was at a higher level than the average diagonal of
290 the matrix (Figure 1). Use of the MAF threshold to make $_{8MAF}$ for $_{22}^{8MAF}$ gave lower
291 average diagonal values than those in . In constructing the $_{8MAF}$ matrix, we deleted the
292 low MAF markers to omit markers with highly uncertain or erroneous AF estimates. This,
293 however, may lead to deleting nearby markers and accepting more markers from certain
294 regions of the genome, particularly if a MAF threshold value higher than 5% is used.
295 Consequently, AF from various MF may become more similar. For example, two breeds may
296 differ due to more intense selection in one of the breeds, leading to the MAF criterion
297 favoring unselected or highly polymorphic markers clustered in certain regions of the
298 genome. Consequently, the matrix may show inflated covariances between the MF of these
299 breeds. Linkage Disequilibrium (**LD**) criteria, in which markers are chosen to minimize LD,
300 is an alternative approach to SNP pruning (Hill and Robertson, 1968). Patterns of LD are

widely used in marker data quality control and in the analysis of population history for various species (Porto-Neto et al., 2014; Makina et al., 2015; Cañas-Álvarez et al., 2016). Multiple studies have shown persistence in LD levels of various breeds and populations (de Roos, 2008; Xu et al., 2019), making LD a potential tool for marker selection.

ssGBLUP estimation & validation results

The correction factor used to calculate the variance of breeding values in base population descending from metafounders (^{2}) in the GEBV calculations for $ssGBLUP_{8}$ and $ssGBLUP_{8MAF}$ was 0.72 and 0.77, respectively. Averages and standard deviations of bull GEBV by birth year are shown in Figures 2 and 3 and the average cow GEBV are shown in Figure 4. We centered the average GEBV trends of cows and bulls, so that the mean GEBV of animals born in 2009 equaled zero. Average bull GEBV in Figure 2 had a similar shape in all the models. The SD level in Figure 3 for bulls born in 2012–2014 was 20 kg (3%) higher in the MF models than in the UPG models. Average cow GEBV by birth year had a similar shape in all models (Figure 4).

Validation test statistics for the approaches are shown in Table 5. Regression coefficients ($\mathbf{b_1}$) were generally slightly higher using MF than UPG. In the bull validation set, we obtained similar adjusted model reliability by $ssGBLUP_{8UPG}$, $ssGBLUP_{8}$, and $ssGBLUP_{8MAF}$, and the gain was 0.04 in comparison to $ssGBLUP_{236UPG}$. In the cow validation set, the validation reliabilities using MF were 0.01 higher than achieved by the UPG models. To exclude pre-selection bias, we conducted the validation tests for bulls also using DYD computed from $ssGBLUP_{236UPG}$. The adjusted model reliabilities did not change from those in Table 5.

Genetic trends in GEBV from the UPG and MF models had a similar shape, showing no effect of the alternative group or founder definitions. We assumed that the inadequate

definition of groups would reduce the genetic trend estimate (Tsuruta et al., 2014) but this was not observed. Each of the bulls included in the yearly means in Figures 2 and 3 had at least 10 daughters and, therefore, may be less affected by MF. Perhaps ssGBLUP predictions where most of the sires are genotyped are robust against the definition of UPG or MF. Meyer and Tier (2018) reported a slightly higher estimated genetic trend with the MF approach compared to ssGBLUP without groups. However, females were the most often genotyped group in their data. Also, the SDs of the GEBV were fairly similar between all evaluations (Figure 3). The unstandardized genetic levels in the MF models were at a higher level compared to the UPG models. This difference did not affect the animal rankings by GEBV but indicate that the models defined base populations differently. We observed a high correlation of bull GEBVs between the MF model and the original 236 UPG model (0.972), while correlation of GEBVs between the MF model and the 8 UPG model was much lower (0.931; correlations not given in Tables).

We used pedigree-based UPG in the ssGBLUP model via incomplete QP transformation (Quaas and Pollak, 1981), i.e. QP transformation for $^{-1}$ instead of $^{-1}$. In case of a multi-breed structure, i.e. for the joint Nordic (Denmark, Finland, Sweden) RDC genetic evaluation, Matilainen et al. (2018) proposed to use QP transformation in $^{-1}$ (Misztal et al., 2013). Bradford et al. (2019) observed that the incomplete QP transformation in ssGBLUP may be applied successfully by accounting for $^{-1}$ only, when a purebred population is analyzed. The MF approach used in this study could be a smooth way to implement the ssGBLUP model for the joint Nordic evaluation.

Estimation of allele frequencies

Defining the base population is the greatest challenge in the MF approach. We focused on two issues: the number of MF and the genetic change in time. Simply replacing current UPG

348 by MF is often impossible in genetic evaluations of large commercial populations, which
349 have many UPG and animals with missing parents. We combined all UPG by breed and split
350 the RDC-based UPG by decade to form eight MF. For the HOL and OTHER breeds, the
351 limited number of animals and absence of phenotypic data were the key reasons for using
352 only one MF per breed. By using multiple MF in RDC, we could account for a possible
353 change in AF with time.

354 Base population AF for the MF are needed to calculate the matrix. Garcia-Baccino et al.
355 (2017) presented three approaches for estimating base population AF to be used for
356 populations with crossbreed animals. All of these methods use genotypes and a pedigree
357 relationship matrix or matrices. We used the genetics group model utilizing GLS. An
358 alternative GLS approach allows differences between gene content variances across breeds
359 and relies on a multi-breed model presented in Garcia-Cortes et al. (2006). All the pedigree-
360 based approaches only need the pedigree of ancestors to the genotyped animals, and the base
361 population groups are defined by MF through pedigree information. However, the
362 unbalanced distribution of genotyped animals to UPG or MF in the full pedigree affects all
363 base population AF estimation methods that rely on the pedigree relationship matrix.

364 In our study, a major part of the genotyped animals (75%) contributed to the oldest RDC
365 group (RDC < 1971) when the full pedigree was used, although most of the genotyped
366 animals (90.6%) were born after 2000. Thus, the contribution gained from genotypes of
367 animals born in 2000–2016 to the recent year groups would be small and would depend on
368 pedigree incompleteness. Consequently, the base population AF of the oldest RDC groups
369 would be well estimated with, possibly, a small influence from young animal genotypes. To
370 solve these issues in the base population AF estimation for the MF, we limited the length of

the pedigree of genotyped animals by only accepting ungenotyped animals with genotyped offspring.

In our study, we calculated the base population AF of HOL and the other breeds group using the ancestor structure of genotyped RDC animals only. We tested the applicability of the chosen GLS approach by estimating an additional matrix ($_{\&}$, Table 6). The matrix was calculated using HOL AF (Koivula 2019, personal communication). We estimated these AF with HOL breed genotypes and the pedigree used in Koivula et al. (2018). The estimated $_{RDC\&HOL}$ was compared with the presented $_{8}$ and $_{8MAF}$ matrices (Table 1), which were only based on genotyped RDC animals. The closeness of the average diagonal values in the HOL MF of $_{RDC\&HOL}$ (0.615), $_{8}$ (0.661), and $_{8MAF}$ (0.593) suggest that we were able to estimate the matrices fairly well without including the pure HOL population genotypes. In addition, the MAF-based marker selection gave the closest value to the HOL genotypes-derived value. Using the truncated pedigree is one possible reason for the good estimation of HOL AF using RDC data. The aim of the pedigree truncation was to distribute available genotypes evenly across MF. Pruning the pedigree appeared to solve two important problems: unequal distribution of genotyped animals across MF and the mixture of AF breed groups.

Off-diagonal elements of the matrix suggested fairly high similarity between all founder groups. We tested a matrix where the off-diagonal elements were half of those in the estimated matrix (results not shown). This half-reduced off-diagonal element matrix nearly gave the same GEBVs solutions, with a correlation of 0.998. Thus, for this data set, the MF-based ssGBLUP evaluation does not seem to be very sensitive to the off-diagonal element values in the matrix. Further work is needed to ascertain that this can be generalized to data sets with more genotyped animals and different population structure.

We observed differences in the matrix depending on the set of markers used to estimate the matrix. When markers were required to have an MAF above a certain limit, values in the matrix were lower than when all the markers were used. This is to be expected because the matrix is estimated by the variance of AF and the MAF threshold reduced range of marker AF is used to calculate the variance. The case is similar to that in Chen et al. (2011) where increasing the MAF threshold in the marker selection decreased the values of (off-)diagonal elements in the genomic relationship matrix. The matrix is a function of the chosen MAF threshold as a consequence of the marker selection. We must therefore be careful when making interpretations of values in the estimated matrix. For example, the MAF threshold was applied to all of the RDC-based MF, but the set of selected markers will change if the HOL animals have genotypes.

The pedigree pruning approach allowed estimation of base population AF for the breed groups despite all the genotyped animals being from the RDC breeding program. Still, it is impossible to model AF changes in base populations and MF before the first genotyped parent generations. One possibility is to assume that the AF changes have continuity and that the changes can also be extrapolated to early years before the genotyping began. Then the variance structures of in the observed base populations, i.e. parents of genotyped animals, could be extended to describe variances of unobservable MF using covariance functions (Kirkpatrick et al. 1994) with appropriate breeds and birth years.

CONCLUSIONS

We tested the metafounder approach on RDC data with a complicated multi-breed structure. The original 236 UPG were replaced by eight MF and tested in ssGBLUP evaluation. Use of MF increased correlation between elements of the pedigree and genomic relationship

matrices. Introduction of MAF-based marker selection before computing the matrix for the MF gave $_{22}^{8MAF}$ an advantage over the original $_{22}^{8}$ in correlations with elements of the genomic relationship matrix. The reduction of UPG groups from 236 to eight reduced the inflation in the predictions and increased validation accuracy. The GEBVs from models with eight MF gave almost the same validation results and genetic trends as the eight UPG. Future development should focus on ways to increase the number of MF closer to the number of UPG.

ACKNOWLEDGEMENTS

We acknowledge Viking Genetics (Randers, Denmark) and Nordic Cattle Genetic Evaluation (Aarhus, Denmark) for providing the genotype data and participating in project financing.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752. https://doi.org/10.3168/jds.2009-2730.

Arnold, J. W., J. K. Bertrand , and L. L. Benyshek. 1992. Animal model for genetic evaluation of multi-breed data. J. Anim. Sci. 70:3322–3332. https://doi.org/10.2527/1992.70113322x.

Bradford, H. L., Y. Masuda, P. M. VanRaden, A. Legarra, and I. Misztal. 2019. Modeling missing pedigree in single-step genomic BLUP. J. Dairy Sci. 102:2336–2346. https://doi.org/10.3168/jds.2018-15434.

Cañas-Álvarez, J. J., E. F. Mouresan, L. Varona, C. Díaz, A. Molina, J. A. Baro, J. Altarriba, M. J. Carabaño, J. Casellas, and J. Piedrafita. 2016. Linkage disequilibrium, persistence of phase, and effective population size in Spanish local beef cattle breeds assessed through a high-density single nucleotide polymorphism chip. J. Anim. Sci. 94:2779–2788. https://doi.org/10.2527/jas.2016-0425.

Chen, C.Y., I. Misztal, I. Aguilar, A. Legarra, and W. M. Muir. 2011. Effect of different genomic relationship matrices on accuracy and scale. J Anim. Sci. 89:2673-2679. doi:10.2527/jas.2010-3555

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2. https://doi.org/10.1186/1297-9686-42-2

Christensen, O. F., P. Madsen, B. Nielsen, and T. Ostersen. 2012. Single-step methods for genomic evaluation in pigs. Animal. 6:1565:1571. https://doi.org/10.1017/S1751731112000742 .

461 Christensen, O. F., A. Legarra, M. S. Lund, and G. Su. 2015. Genetic evaluation for three-
462 way crossbreeding. Genet. Sel. Evol. 47:98. https://doi.org/10.1186/s12711-015-0177-
463 6.

464

465 de Roos, A. P., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium
466 and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. Genetics.
467 179:1503–1512. https://doi.org/10.1534/genetics.107.084301.

468

469 Gao, H., M. Koivula, J. Jensen, I. Strandén, P. Madsen, T. Pitkänen, G. P. Aamand, and E. A.
470 Mäntysaari. 2018. Short communication: Genomic prediction using different single-
471 step methods in the Finnish red dairy cattle population. J. Dairy Sci. 101:10082–10088.
472 https://doi.org/10.3168/jds.2018-14913.

473

474 Garcia-Baccino, C. A., A. Legarra, O. F. Christensen, I. Misztal, I. Pocrnic, Z. G. Vitezica,
475 and R. J. Cantet. 2017. Metafounders are related to Fst fixation indices and reduce bias
476 in single-step genomic evaluations. Genet. Sel. Evol. 49:34.
477 https://doi.org/10.1186/s12711-017-0309-2.

478

479 Garcia-Cortes, L. A., and M. A. Toro. 2006. Multibreed analysis by splitting the breeding
480 values. Genet. Sel. Evol. 38:601-15. https://doi.org/10.1051/gse:2006024.

481

Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. J. Dairy Sci. 93:1243-1252. https://doi.org/10.3168/jds.2009-2619.

Hill, W. G., and A. Robertson. 1968. Linkage disequilibrium in finite populations. Theor. Appl. Genet. 38:226–31. https://doi.org/10.1007/BF01245622.

Kirkpatrick, M., W. G. Hill, and R. Thompson. 1994. Estimating the covariance structure of traits during growth and ageing, illustrated with lactation in dairy cattle. Genet Res. 64:57–69.

Koivula, M., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2018. Comparison of ssGBLUP and ssGTBLUP using Nordic Holstein TD data. Processing of the World Congress on Genetics Applied to Livestock Production. 11:445.

Legarra A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663. https://doi.org/10.3168/jds.2009-2061.

Legarra, A., O. F. Christensen, Z. G. Vitezica, I. Aguilar, and I. Misztal. 2015. Ancestral relationships using metafounders: finite ancestral populations and across population relationships. Genetics. 200:455–468. https://doi.org/10.1534/genetics.115.177014.

Lidauer, M. H., J. Pösö, J. Pedersan, J. Lassen, P. Madsen, E. A. Mäntysaari, U. S. Nielsen, J.-A. Eriksson, K. Johanson, T. Pitkänen, I. Strandén, and G. P. Aamand. 2015. Across-country test-day model evaluations for Holstein, Nordic Red Cattle, and Jersey. J. Dairy Sci. 98:1296–1309. https://doi.org/10.3168/jds.2014-8307.

Makina, S. O., J. F. Taylor, E. Van Marle-Köster, F. C. Muchadeyi, M. L. Makgahlela, M. D. MacNeil, and A. Maiwashe. 2015. Extent of Linkage Disequilibrium and Effective Population Size in Four South African Sanga Cattle Breeds. Front. Genet. 6:337. https://doi.org/10.3389/fgene.2015.00337.

Masuda Y., S. Tsuruta, E. Nicolazzi and I. Misztal. 2019. Single-step GBLUP including more than 2 million genotypes with missing pedigrees for production traits in US Holstein. Interbull Open Meeting: 22-23 June, Cincinnati, Ohio, USA. https://interbull.org/static/web/10_30_Masuda_final.pdf

Matilainen, K., I. Strandén, G. P. Aamand, and E. A. Mäntysaari. 2018. Single step genomic evaluation for female fertility in Nordic Red dairy cattle. J. Anim. Breed. Genet. 135:337-348. https://doi.org/10.1111/jbg.12353.

McPeek, M. S., W. Xiaodong, and C. Ober. 2004. Best Linear Unbiased Allele-Frequency Estimation in Complex Pedigrees. Biometrics 60:359–67. https://doi.org/10.1111/j.0006-341X.2004.00180.x.

Meyer, K., B. Tier, and A. Swan, 2018. Estimates of genetic trend for single-step genomic evaluations. Genet. Sel. Evol. 50:39. https://doi.org/10.1186/s12711-018-0410-1.

Misztal, I., Z. G. Vitezica, A. Legarra, I. Aguilar, and A. A. Swan. 2013. Unknown-parent groups in single-step genomic evaluation.. J. Anim. Breed. Genet. 130:252–258. https://doi.org/10.1111/jbg.12025

Mäntysaari, E. A., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. Interbull Bull.41:17–22.

Porto-Neto, L. R., J. W. Kijas, and A. Reverter. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. 2014. Genet. Sel. Evol. 46:22. https://doi.org/10.1186/1297-9686-46-22.

Quaas, R. L., and E. J. Pollak. 1981. Modified equations for sire models with groups. J. Dairy Sci. 64:1868–1872. https://doi.org/10.3168/jds.S0022-0302(81)82778-6.

546 R Development Core Team (2008). R: A language and environment for statistical computing.
547 R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL
548 http://www.R-project.org.

549

550 Sponenberg, D. P., and D. E. Bixby. 2007. Managing Breeds for a Secure Future: Strategies
551 for Breeders and Breed Associations. American Livestock Breeds Conservancy.
552 Pittsboro, NC.

553

554 Strandén, I., and M. Lidauer. 1999. Solving large mixed models using preconditioned
555 conjugate gradient iteration. J. Dairy Sci. 82:2779–2787.
556 https://doi.org/10.3168/jds.S0022-0302(99)75535-9.

557

558 Strandén, I., M. Lidauer, E. A. Mäntysaari, and J. Pösö. 2000. Calculation of Interbull
559 weighting factors for the Finnish test day model. Interbull Bull. 26:78–79.

560

561 Strandén, I., and K. Vuori. 2006. RelaX2: pedigree analysis program. Proc. 8th WCGALP,
562 Belo Horizonte, Brazil.

563

564 Strandén, I., K. Matilainen, G. P. Aamand, and E. A. Mäntysaari. 2017. Solving efficiently
565 large single-step genomic best linear unbiased prediction models. J. Anim. Breed.
566 Genet. 134:264–274. https://doi:.org/10.1111/jbg.12257.

Strandén, I., E.A. Mäntysaari. 2018. HGinv Program. Natural Resources Institute Finland (LUKE).

Strandén, I., E.A. Mäntysaari. 2019. Bpop Program. Natural Resources Institute Finland (LUKE).

Theron, H. E., F. H. J. Kanfer, and L. Rautenbach. 2002. The effect of phantom parent groups on genetic trend estimation. S. Afr. J. Anim. Sci. 32: 130–135. https://doi.org/10.4314/sajas.v32i2.3755.

Thompson, R. 1979. Sire evaluation. Biometrics. 33:497–504. https://doi.org/10.2307/2529955.

Tsuruta, S., I. I. Misztal, D. Lourenco, and T. Lawlor. 2014. Assigning unknown parent groups to reduce bias in genomic evaluations of final score in US Holsteins. J. Dairy Sci. 97:5814-5821. https://doi.org/10.3168/jds.2013-7821.

van Grevenhof, E. M., J. Vandenplas, M. P. L. Calus. 2019. Genomic prediction for crossbred performance using metafounders, J. Anim. Sci. 97:548–558. https://doi.org/10.1093/jas/sky433.

588

589 VanRaden P. M. 1992. Accounting for inbreeding and crossbreeding in genetic evaluation of
590 large populations. J. Dairy Sci. 75:3136–3144. https://doi.org/10.3168/jds.S0022-
591 0302(92)78077-1.

592

593 VanRaden P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci.
594 91:4414–4423. https://doi.org/10.3168/jds.2007-0980.

595

596 Vitezica Z., I. Aguilar, I. Misztal, and A. Legarra. 2011. Bias in genomic predictions for
597 populations under selection. Genet. Res. 93:357–366.
598 https://doi.org/10.1017/S001667231100022X.

599

600 Wiggans, G. R., T. S. Sonstegard, P. M. VanRaden, L. K. Matukumalli, R. D. Schnabel, J. F.
601 Taylor, F. S. Schenkel, and C. P. Van Tassell. 2009. Selection of single-nucleotide
602 polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in
603 the United States and Canada. J. Dairy Sci. 92:3431–3436.
604 https://doi.org/10.3168/jds.2008-1758.

605

606 Xiang, T., O. F. Christensen, and A. Legarra. 2017. Genomic evaluation for crossbred
607 performance in a single-step approach with metafounders. J. Anim. Sci. 95: 1472–1480.
608 https://doi.org/10.2527/jas.2016.1155.

609

610 Xu, L., B. Zhu, Z. Wang, L. Xu, Y. Liu, Y. Chen, L. Zhang, X. Gao, H. Gao, S. Zhang, L.
611 Xu, J. Li. 2019. Evaluation of Linkage Disequilibrium, Effective Population Size and
612 Haplotype Block Structure in Chinese Cattle. Animals (Basel). 9:83.
613 https://doi.org/10.3390/ani9030083.

614

615 **Table 1.** Estimated $_8$ (lower) and $_{8MAF}$ (upper) triangle for the metafounders. The diagonal

616 includes diagonals (i.e. self-relationships of metafounders) of $_8$ (in brackets) and $_{8MAF}$.

	RDC[1] <1970	RDC[1] 1971–1980	RDC[1] 1981–1990	RDC[1] 1991–2000	RDC[1] 2001–2010	RDC[1] 2011–2016	HOL[1]	OTHER[1]
RDC[1] <1970	0.618 (0.719)	0.555	0.563	0.563	0.566	0.566	0.471	0.453
RDC[1] 1971–1980	0.659	0.569 (0.670)	0.566	0.561	0.564	0.562	0.473	0.454
RDC[1] 1981–1990	0.668	0.670	0.609 (0.710)	0.588	0.589	0.585	0.473	0.452
RDC[1] 1991–2000	0.667	0.664	0.690	0.587 (0.689)	0.585	0.583	0.473	0.455
RDC[1] 2001–2010	0.671	0.667	0.692	0.688	0.598 (0.701)	0.597	0.474	0.452
RDC[1] 2011–2016	0.671	0.666	0.688	0.686	0.699	0.603 (0.705)	0.474	0.453
HOL[1]	0.563	0.564	0.564	0.564	0.566	0.566	0.593 (0.661)	0.479
OTHER[1]	0.544	0.544	0.544	0.545	0.544	0.545	0.552	0.740 (0.797)

617 [1]Red dairy cattle (RDC) has been divided into metafounders by birth year, Holstein (HOL)

618 cattle has one metafounder, and the other breeds (OTHER) have been combined into one

619 metafounder.

620

621

622 **Table 2.** Inbreeding coefficients of metafounders calculated using $_8$ and $_{8MAF}$.

Groups[1]	Γ_8	Γ_{8MAF}.
RDC <1970	-0.28	-0.38
RDC 1971–1980	-0.33	-0.43
RDC 1981–1990	-0.29	-0.39
RDC 1991–2000	-0.31	-0.41
RDC 2001–2010	-0.29	-0.40
RDC 2011–2016	-0.29	-0.39
HOL	-0.34	-0.40
OTHER	-0.34	-0.26

623 [1]Red dairy cattle (RDC) has been divided into metafounders by birth year, Holstein (HOL)
624 cattle has one metafounder, and the other breeds (OTHER) have been combined into one
625 metafounder.

626

627 **Table 3.** Correlation of diagonal (upper triangle) and off-diagonal (lower triangle) elements

628 of $_{22}$, , , $^{8}_{22}$, and $^{8MAF}_{22}$.

	$_{22}$	$^{8}_{22}$	$^{8MAF}_{22}$		
$_{22}$	1	0.81	0.84	0.66	0.53
$^{8}_{22}$	0.89	1	0.99	0.76	0.33
$^{8MAF}_{22}$	0.92	0.99	1	0.76	0.37
	0.83	0.91	0.91	1	0.70
	0.89	0.86	0.88	0.88	1

629

630 **Table 4.** Mean, minimum (Min), and maximum (Max) element values of $_{22}$, ,
631 , $^{8}_{22}$, and $^{8MAF}_{22}$ from diagonal and off-diagonal.

Elements	Matrix	Mean	Min	Max
Diagonal	$_{22}$	1.02	1.00	1.29
		1.31	1.24	1.48
		1.01	0.91	1.30
	$^{8}_{22}$	1.35	1.27	1.51
	$^{8MAF}_{22}$	1.31	1.23	1.50
Off-diagonal	$_{22}$	0.07	0.06	0.81
		0.63	0.47	1.29
		0.05	-0.11	0.99
	$^{8}_{22}$	0.72	0.54	1.22
	$^{8MAF}_{22}$	0.62	0.45	1.16

632

633 **Table 5.** GEBV validation test regression coefficients and validation reliabilities of single-
634 step GBLUP GEBVs for genotyped bulls and cows.

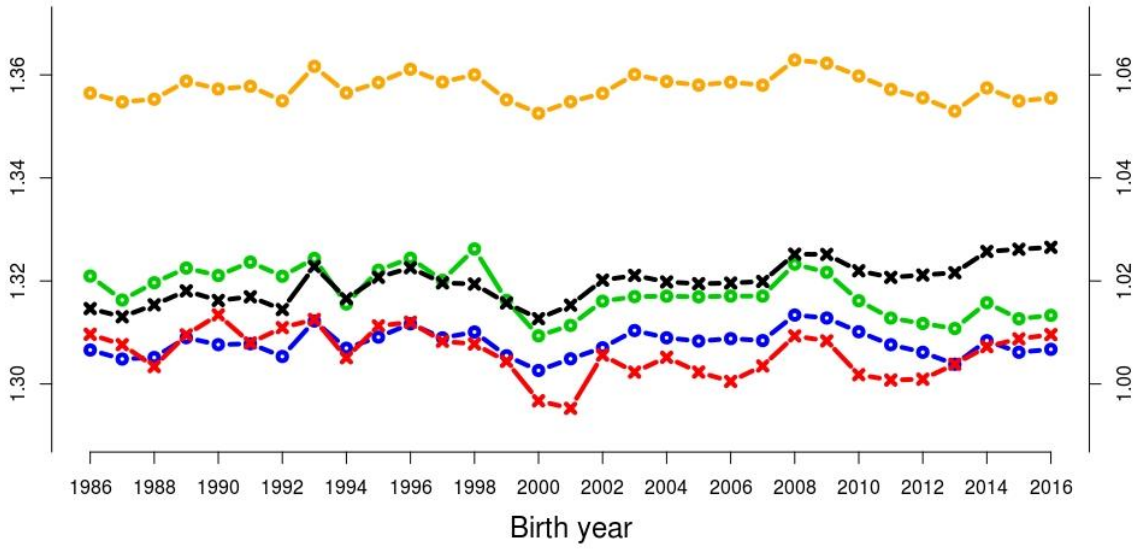
Validation set	Model[1]	$_0$	SE	$_1$[2]	SE	R^2 [3]	R^2 [3]
Bulls	ssGBLUP$_{236UPG}$	70	16	0.61	0.06	0.23	0.27
	ssGBLUP$_{8UPG}$	18	16	0.73	0.06	0.26	0.31
	ssGBLUP$_{8}$	-22	22	0.72	0.06	0.26	0.31
	ssGBLUP$_{8MAF}$	-27	23	0.73	0.06	0.26	0.31
Cows	ssGBLUP$_{236UPG}$	118	9	0.89	0.03	0.16	0.36
	ssGBLUP$_{8UPG}$	150	8	0.89	0.03	0.16	0.36
	ssGBLUP$_{8}$	12	13	0.90	0.03	0.16	0.37
	ssGBLUP$_{8MAF}$	-0.2	13	0.93	0.04	0.16	0.37

635 [1]Model ssGBLUP$_{236UPG}$ (ssGBLUP$_{8UPG}$) had 236 (8) unknown parent groups; ssGBLUP$_{8}$ had 8
636 metafounders with the metafounder matrix calculated using all markers; ssGBLUP$_{8MAF}$ used
637 markers with a minor allele frequency ≥ 0.05 in the metafounder matrix calculation.

638 [2] Regression coefficient $_1$ in equation = $_0$ + $_1$ for the bulls has been multiplied by
639 2.

640 [3] R^2 is the coefficient of determination from the validation regression, R^2 is adjusted by the average
641 reliability of phenotypes in the validation group.

642



643

644 **Figure 1.** Average diagonal elements of $_{22}$ (black cross), (red cross), (green

645 circles)**,** $^{8}_{22}$ (orange circles), and $^{8MAF}_{22}$ (blue circles) by the birth year of an animal. The

646 left side of the y-axis has a scale for **,** $^{8}_{22}$ and $^{8MAF}_{22}$ and the right side has a scale for

647 $_{22}$ and .

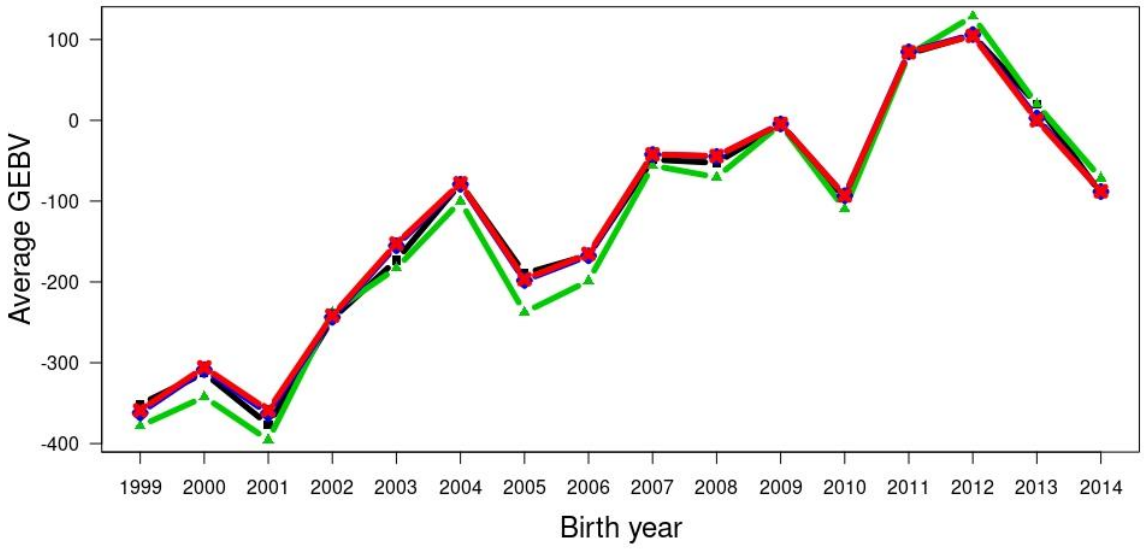
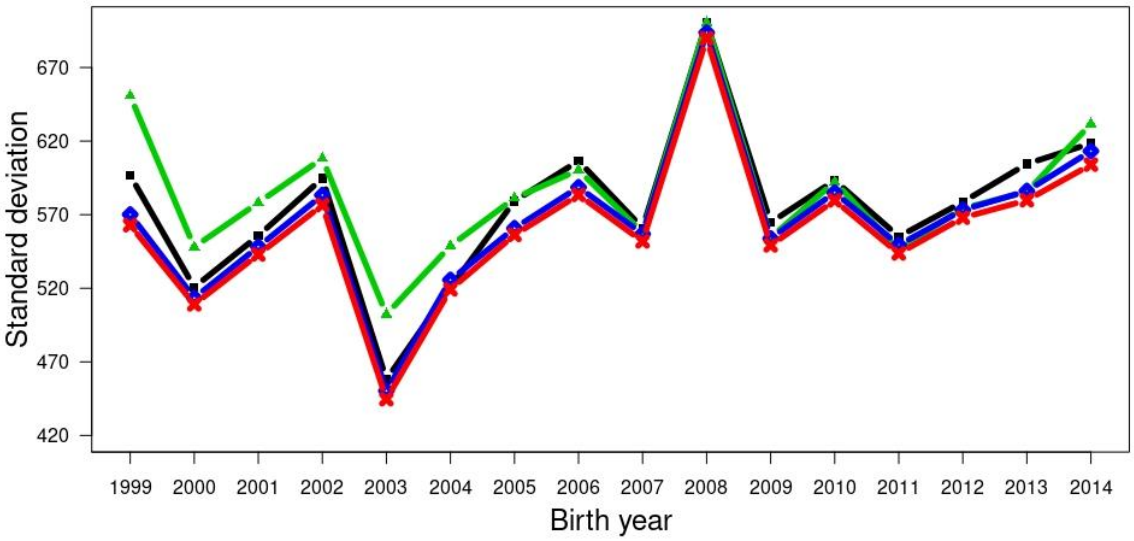


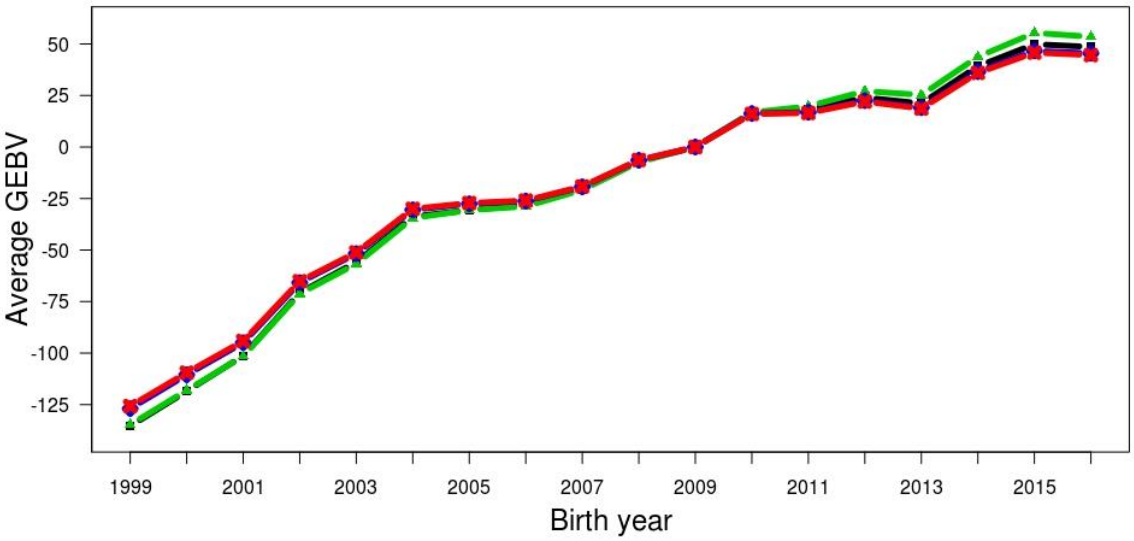
Figure 2. Average genomic breeding value of bulls by birth year in 305-d milk yield (kg). Each bulls had at least 10 daughters. The lines above each other are from the unknown parent group models ssGBLUP$_{236UPG}$ (black square) and ssGBLUP$_{8UPG}$, (green triangle) and from the metafounders models ssGBLUP$_{8}$ (blue diamond) and ssGBLUP$_{8MAF}$ (red cross).



654

655 **Figure 3.** Standard deviation of bull genomic breeding values by birth year in 305-d milk
656 yield, kg. Each bull had at least 10 daughters. Trends are from the unknown parent group
657 models ssGBLUP$_{236UPG}$ (black square) and ssGBLUP$_{8UPG}$, (green triangle) and from the
658 metafounders models ssGBLUP $_{8}$ (blue diamond) and ssGBLUP $_{8MAF}$ (red cross).

659



660

661 **Figure 4.** Average genomic breeding value of cows by birth year in 305-d milk yield (kg).

662 The lines above each other are from the unknown parent group models ssGBLUP$_{236UPG}$

663 (black square) and ssGBLUP$_{8UPG}$, (green triangle) and from the metafounders models

664 ssGBLUP $_{8}$ (blue diamond) and ssGBLUP $_{8MAF}$ (red cross).

665

666

667 Table 6. Gamma matrix created using base population allele frequencies calculated from Red

668 Dairy Cattle (RDC) and Holstein (HOL) cattle genotypes.

	RDC[1] <1970	RDC[1] 1971–1980	RDC[1] 1981–1990	RDC[1] 1991–2000	RDC[1] 2001–2010	RDC[1] 2011–2016	OTHER[1]	HOL[1] <1970	HOL[1] 1970–1980	HOL[1] 1981–1990	HOL[1] 1991–2000	HOL[1] 2001–2010	HOL[1] 2011–2016
RDC[1] <1970	0.825	0.613	0.602	0.604	0.604	0.603	0.536	0.521	0.533	0.524	0.516	0.515	0.512
RDC[1] 1971–1980		0.638	0.629	0.629	0.627	0.622	0.539	0.521	0.539	0.526	0.516	0.515	0.512
RDC[1] 1981–1990			0.665	0.665	0.657	0.648	0.543	0.520	0.538	0.525	0.515	0.514	0.512
RDC[1] 1991–2000				0.670	0.664	0.654	0.543	0.520	0.538	0.525	0.516	0.515	0.512
RDC[1] 2001–2010					0.676	0.668	0.542	0.520	0.538	0.525	0.515	0.515	0.512
RDC[1] 2011–2016						0.666	0.547	0.521	0.539	0.526	0.517	0.516	0.514
OTHER[1]							0.813	0.511	0.525	0.518	0.509	0.507	0.503
HOL[1] <1970								0.581	0.559	0.579	0.586	0.587	0.589
HOL[1] 1970–1980									0.574	0.567	0.562	0.561	0.560
HOL[1] 1981–1990										0.595	0.594	0.595	0.598
HOL[1] 1991–2000											0.613	0.615	0.621
HOL[1] 2001–2010												0.628	0.638
HOL[1] 2011–2016													0.690

669 [1]RDC and HOL cattle have been divided into metafounders by birth year, while the other

670 breeds (OTHER) have been combined into one metafounder.

671